EE

EUROPEAN LABORATORY FOR PARTICLE PHYSICS

CERN - SL DIVISION

CERN SL/95-07 (AP)

Algorithms and procedures used in the orbit correction package COCU (Closed Orbit Correction Utilities)

W. Herr

CERN SL Division, 1211-Geneva 23

Abstract

The closed orbit correction package COCU (Closed Orbit Correction Utilities) was originally written for the orbit correction at SPS and LEP but has now found a wider distribution. In this report a description of the algorithms and procedures implemented in COCU is presented. Emphasis was put on a presentation of the basic principles and not the underlying mathematics or the implementation. The advantages and disadvantages of the various methods are discussed and the data preparation in COCU is described. This report complements the User Guide and the Programmers manual.

Geneva, Switzerland

May 23, 1995

1 Introduction

In this report I shall describe all the algorithms and procedures used in the closed orbit correction package COCU [1, 2] which has originally been developed for SPS and LEP but has now found a wider distribution. This report will not describe how the algorithms are implemented or used and therefore is neither a user guide nor a programmers manual. This information can be found in other reports [3, 4].

It will be tried to explain the basic properties and the principles of the different algorithms and to discuss their advantages and disadvantages. A detailed mathematical treatment will be avoided as this can be found in the appropriate references.

In the first section I shall formulate the problem and give some basic properties of orbit correction algorithms. In the following sections the global and local orbit correction algorithms are described and in the last section the data preparation is presented as it forms an important part of the package.

2 Problem of closed orbit correction

2.1 The purpose

The correction of the closed orbit is a problem which is common to all circular accelerators and proves to be an important issue for the performance of the machine.

At LEP is was found that a good control of the orbit is of great importance to achieve good luminosities and lifetime [5, 6].

Apart from ensuring that the beam remains inside the physical aperture of the beam pipe, the closed orbit influences the dynamics of the beam through various effects whose importance may vary between different accelerators and the type of particles accelerated (i.e. hadron or leptons). Examples for such effects are:

- Large closed orbit decreases the dynamic aperture
- Good orbit required for transverse polarization in LEP
- Excitation of unwanted dispersion and change of damping
- Bad orbit induces background in experiments
- Luminosity optimization is affected

Furthermore, a good closed orbit is often a prerequisite for other corrections and manipulations, e.g. the coupling correction and other fine tuning such as separator adjustment.

2.2 Formulation of the problem

The beam position with respect to the ideal orbit is measured in m locations around the ring and the resulting measured closed orbit can be written as a vector $\vec{x_m}$ of length m:

$$\vec{x}_m = (x_1, ..., x_m)^t$$

Assuming another vector of length n containing the strengths of closed orbit correction dipoles:

$$\vec{y} = (y_1, ..., y_n)$$

the effect of these correctors on the closed orbit can be written as:

$$\vec{x}_r = A \cdot \vec{y} \tag{1}$$

resulting in a closed orbit $\vec{x}_r$. Here A is the response matrix of size [m x n] where the j^{th} column describes the effect of the j^{th} corrector with a unit excitation on the m beam positions. A single element a_{ij} of matrix A describes therefore the effect of the j^{th} corrector on the i^{th} monitor and follows the equation:

$$a_{ij} = \frac{\sqrt{\beta_i \cdot \beta_j}}{2sin(\pi Q)} \cdot cos(|\mu_i - \mu_j| - \pi Q) \tag{2}$$

where Q is the overall tune, μ_i and μ_j their betatron phases and β_i and β_j their β-functions.

The task of an orbit correction is to find a set of correctors $\vec{y}$ to satisfy:

$$\vec{x}_t = A \cdot \vec{y} + \vec{x}_m \tag{3}$$

where $\vec{x}_t$ is the target orbit. Normally this target is an orbit where all beam positions are zero, i.e. the beam is centred and equation (3) becomes:

$$\vec{x}_t \equiv 0 \quad \rightarrow \quad \vec{x}_m = -A \cdot \vec{y} \tag{4}$$

This is not a necessary condition and a priori the target orbit can be chosen freely to solve the specific problem. In LEP, the orbit is frequently corrected towards a so-called "golden orbit" which was found empirically to give good luminosity and background conditions. From now on, the orbit to be corrected is the difference between the measured and the target orbit.

The problem as described here is one-dimensional, i.e. the algorithms are applied in one plane only and coupling effects are not taken into account. An iterative procedure may be required to obtain a good orbit in both planes.

2.3 Boundary conditions and errors

The quality of a closed orbit correction strongly depends on the boundary conditions imposed by the layout and the operation of the machine.

The precision and reliability of the beam position measurement is naturally of vital importance as a correction can only be as good as the measurement. The number of monitors and correctors per phase advance will define the sampling of the closed orbit measurement and its correction. Any undersampling of the measurement will lead to an unknown orbit in a part of the machine and an insufficient number of correctors could make a good correction impossible, even for a perfect measurement.

The precision of the correctors and their maximum strengths may further limit the quality of the correction.

A procedure to correct the closed orbit has to take into account these possible problems and perform the necessary precautions prior to the application of an algorithm and possibly by postprocessing the results.

As these techniques form a vital part of the correction procedure, the most important of them are discussed in a later section.

3 Global closed orbit correction procedures

3.1 Quality of orbit correction and norm

In general, the number of measurements (m) and the number of correctors (n) are not the same and the system of linear equation (3) is either overdetermined (m > n) or underdetermined (m < n). The latter is however very unlikely and normally the number of monitors exceeds the number of correctors.

In this case the system of equation (3) cannot be solved exactly and an approximate solution is wanted:

$$\vec{x}_a = A \cdot \vec{y} + \vec{x}_m \tag{5}$$

where $\vec{x}_a$ is close to $\vec{x}_t$ but not equal.

The function

$$S = \| \vec{x}_t - \vec{x}_a \|_{l_n} \tag{6}$$

measures the distance from the desired orbit using the norm l_n as an estimator.

The closed orbit correction algorithms have to find a set of $\vec{y}$ to satisfy:

$$\vec{y} \rightarrow S \quad = \quad \| \vec{x}_t - \vec{x}_a \|_{l_n} \quad = \quad minimum \tag{7}$$

The choice of the norm is a priori free and can be chosen appropriately for the problem one wants to solve. In most cases the least squares or quadratic norm l_2 is used and therefore a least squares minimization is applied. Other commonly used norms are the l_1 norm (used in SIMPLEX) which treats errors independently of their amplitude and the Tchebyshev norm l_∞ which allows to control the maximum beam position. The choice of the norm defines the algorithm to solve the problem (7) and in the following the norm is always specified when an algorithm is discussed.

To quantify the quality of an orbit usually the "peak-to-peak" value, i.e. the distance between the two extreme amplitudes, and the r.m.s. value with respect to the average orbit are specified.

3.2 Orbit correction strategies

Depending of the problem to solve and the configuration of the closed orbit different strategies may be applied and have to be available. Very often the closed orbit deviation from the ideal orbit is dominated by a few strong and localized distortions, in particular at the beginning of a correction procedure when the orbit deviations are rather large. In that case it is useful to first identify the few dominant defects and correct for these.

We therefore need methods to find such dominant kicks fast and reliably. These methods are usually classified as "best kick" methods.

Furthermore, when the orbit is large and the beam travels off centred through quadrupoles, sextupoles etc., second order effects modify the closed orbit and an iterative procedure is required. One would usually reduce the orbit deviations to "reasonable" small values with a few correctors before a correction towards the best possible orbit is attempted with a large number of correctors. Therefore the number of correctors used for each correction in the procedure is usually increasing as the orbit becomes better.

3.3 Best kick orbit correction methods

3.3.1 MICADO ALGORITHM

The most frequently used method at LEP is the MICADO (MINimisation des Carrés de Distortion d'Orbit) minimization [7] which uses a least squares norm but the strategy to find the solution is quite different from a standard least squares minimization. It can be shown that for a least squares minimization the minimum norm can be rewritten as

$$S^2_{min} = \vec{y} \cdot A^t \cdot \vec{x}_m + \vec{x}^t_m \cdot \vec{x}_m \qquad (8)$$

The matrix A^t is the transpose of the response matrix A. The vector $A^t \cdot \vec{x}_m$ is the correlation product of the effect of a given corrector on the measured orbit. This correlation product can be used for an iterative minimization procedure. At the beginning of a correction, each corrector is tested and that with the highest correlation is kept as the most efficient corrector and its strength is calculated.

In each further iteration, this correlation is established between the remaining correctors and the residual orbit after the previous step. The corrector with the next highest correlation is included in the list of correctors used for the correction and the strengths of all previously selected correctors are recalculated. The strength of a given corrector may therefore change when further correctors are included.

This procedure is performed by means of orthogonal transformations and the mathematical details are described in [7].

A particular feature of the procedure described above is that once a corrector is selected, it will remain in the list, although its strength may change after more iterations. Therefore, having selected k correctors after k iterations, this is not necessarily the best combination of k correctors to solve the problem.

However, in the present implementation of the algorithm in COCU, it is possible to suggest the first corrector to the MICADO algorithm to provide some "guidance" for the procedure.

3.3.2 BEST COMBINATION - MINIMO ALGORITHM

Another type of minimization with a given number of correctors is implemented as MINIMO. This type of correction tries to overcome the problem mentioned above that with k chosen correctors not the best subset for that number of iterations is chosen. The procedure in MINIMO is rather straightforward and simple. Assuming a total of n correctors available and k correctors desired, all possible combinations are tested.

For each combination the problem is formulated as:

$$minimize: \qquad \vec{y}_{sub} \cdot A_{sub} + \vec{x}_m \qquad (9)$$

where A_{sub} is now a subset of the response matrix A for only the k correctors under consideration (A reduced from [m x n] to a [m x k] matrix). The solution vector $\vec{y}_{sub}$ is now calculated with a standard least squares minimization:

$$\vec{y}_{sub} = (A_{sub} \cdot A^t_{sub})^{-1} \cdot A^t_{sub} \cdot \vec{x}_m \qquad (10)$$

The combination with the smallest residual orbit is retained as the best combination.

Although this method would always give the best combination of correctors, it has some severe disadvantages. For example, for a large number of correctors available the number of combinations to be tested becomes very large and the required computing time is unacceptable. In table Tab.1 I have illustrated this by giving the number of combinations and the CPU time required on a high performance workstation (HP 735) for the case of LEP with 284 correctors. It can clearly be seen that already with a small number of iterations the required computing time quickly becomes impractical.

Table 1: Time necessary for MINIMO correction (on an idle HP-735, 99 MHz)

Iterations (correctors):	Number of combinations:	CPU-time required:
1	284	0.10 sec
2	40186	21.0 sec
3	$\approx 3.8 \cdot 10^{6}$	$\approx$ 50 min
5	$\approx 1.5 \cdot 10^{10}$	$\approx$ 120 days
10	$\approx 8.0 \cdot 10^{17}$	$\approx 1.7 \cdot 10^{7}$ years

3.3.3 MIN2C

As the MINIMO algorithm is too time consuming and the MICADO algorithm has some disadvantages, another algorithm has been added which is some type of compromise. The MIN2C (MInimization with 2 Correctors) algorithm is equivalent to the MINIMO algorithm but restricted and optimized to find *pairs* of correctors.

It can therefore be particularly useful to identify and correct bumps formed by two magnets. By construction, the MICADO algorithm would find bumps only accidentally when it is run for many iterations (see later for details).

Furthermore, it can either be iterated to find successively pairs of correctors or the result from this algorithm can be used as the first two correctors for the MICADO algorithm.

3.3.4 SIMPLEX

The SIMPLEX algorithm is used to calculate a solution using the l_1 norm for the minimization of:

$$\| \vec{y} \cdot A + \vec{x}_m \| \tag{11}$$

but also allows to subject the algorithm to linear constraints of the type:

$$\vec{y} \cdot B = \vec{b} \tag{12}$$

or inequality constraints of the type:

$$\vec{y} \cdot C \leq \vec{c} \tag{13}$$

where A, B and C are matrices and $\vec{b}$ and $\vec{c}$ vectors. An example for constraints of the type (13) would be the maximum possible correctors strength and such constraints are used in the SPS orbit correction.

The problem (11) - (13) is subjected to a standard minimization procedure using the SIMPLEX method [10].

3.4 Complete least squares minimization

3.4.1 PSINOM ALGORITHM

In case the response matrix is not rank deficient, a global least squares minimization can be attempted.

The linear system of equation:

$$\vec{x}_m = -A \cdot \vec{y} \tag{14}$$

has an approximate solution with minimum least squares of the form:

$$\vec{y} = (A \cdot A^t)^{-1} \cdot A^t \cdot \vec{x}_m \tag{15}$$

very similar to (10), however with the full response matrix using all available correctors.

This algorithm has been implemented as the "PSINOM" mode but it requires a non-singular response matrix. The matrix must therefore be properly conditioned, e.g. with a singular value decomposition as described later.

The result of this algorithm will be a correction using all available correctors. The result is equivalent to a correction with the MICADO algorithm when the number of MICADO iterations equals the number of correctors.

3.5 Trajectory correction and first turn steering

The correction of a trajectory or a closed orbit are equivalent except that the response matrix must reflect the different boundary conditions:

- Closed orbit:
  Orbit is closed around the ring
  Particles have no starting point in the ring
  All correctors affect all measurements
  Closed orbit the same for both charges

- Trajectory:
  Trajectory has no closure condition
  Particles have a starting point in the ring
  Only correctors before a monitor affect a measurement
  Particles of different charge travel different directions and have different trajectories

Provided the necessary modifications to the response matrix are performed, the available algorithms can be used for a trajectory correction.

3.5.1 Preparation for trajectory correction

The element a_{ij} of the response matrix A describing a trajectory is now:

$$a_{ij} = \delta_j \cdot \sqrt{\beta_i \cdot \beta_j} \cdot sin(|\mu_i - \mu_j|) \tag{16}$$

using the same variables as for the closed orbit correction.

Since the trajectory has a natural starting point, the location of the injection has to be specified. In COCU this information is used to scale the theoretical phases to start at this position. This assumes that the particles travel clockwise around the machine.

Furthermore, only those elements a_{ij} are different from zero where the j^{th} corrector has a smaller phase than the i^{th} monitor. A corrector can only have an effect on a monitor if it is located in front of it since a trajectory is only a single pass through the machine. The response matrix has therefore already a "triangular-like" shape.

For particles traveling counterclockwise in the ring, the phases must increase in the direction of the particles, the phases are therefore reflected at the injection position.

A single parameter to COCU will initiate the necessary actions described above.

3.5.2 Use of MICADO algorithm

The MICADO algorithm can be used to "solve" the problem stated above. Any of the other algorithms can be used as well, but especially for a trajectory or first turn correction only one or a very few correctors are normally used and MICADO is always the best and fastest choice.

3.6 Additional orbit manipulation facilities

3.6.1 Harmonic analysis and filtering

To suppress noise from the measurements, a harmonic filtering has been implemented [11].

A harmonic analysis is performed on the input data and if required a new orbit is generated with a limited number of harmonics centred around the nominal tune value.

This allows a noise reduction for a following correction with any of the implemented algorithms.

It can further be used to determine the integer part of the tune by analyzing a difference orbit before and after a single corrector has been applied.

3.6.2 Orbit simulation

The effect of a specified set of correctors $\vec{y}$ on the orbit can be calculated as:

$$\vec{x}_r = A \cdot \vec{y} \tag{17}$$

The orbit $\vec{x}_r$ can either be added to the existing orbit (absolute orbit) or only the increments are saved. The resulting orbit vector can be treated as an output of a normal correction.

In particular the corrector settings are treated as from a normal correction algorithm and can be send to the equipment.

4 Local closed orbit correction procedures

Very often it is required to correct the orbit locally without changing it in the rest of the machine. This may be required to reduce the background in an experiment or to provide a controlled orbit distortion at a specific location.

The two ways implemented in COCU perform this either with local, closed bumps or by a localized least squares minimization with a closure condition outside the short length region.

4.1 Short length corrections

The least squares minimization of a short region can in principle be performed with any of the algorithms implemented.

The closure condition can be enforced either by an explicit closure after the correction has been found using two additional correctors or by changing the weights such as to emphasize a correction in a specific region. The latter method has been implemented together with a MICADO minimization while the first type is used in a dedicated algorithm used for short length correction using a gradient projection method [8].

4.1.1 Use of MICADO algorithm

To force MICADO to perform a localized correction it is enough to change the weights in such a way that the contribution from the desired region determines largely the norm which is minimized. The other part is considered "already good" by the algorithm and MICADO would not find a correction which would strongly deteriorate this part of the ring. As long as the short length region treated is small compared to the rest of the ring, this method works well and reliably [1, 2, 8].

A disadvantage of this method is that the correction is not always *completely* closed as the closure is not done explicitly. The main advantage is the possibility to use a best kick strategy together with the MICADO algorithm also for short length corrections.

It should be mentioned although it is obvious, that in addition to the correctors used for the actual short length correction at least two to three additional correctors must be available to MICADO to "close" the correction. The minimum number of correctors (iterations) used should therefore be never smaller than four.

4.1.2 GRAPE

A different approach was followed for the second short length correction method which is based on a Gradient Projection Method (GPM) which is a steepest descent method and finds a minimum of a function under linear constraints. The problem to be solved is again a minimization of the function:

$$S = \parallel \vec{x}_t - \vec{x}_a \parallel_{l_n} \tag{18}$$

with linear constraints of the type:

$$\vec{y} \cdot B = \vec{b} \tag{19}$$

$$\vec{y} \cdot C \leq \vec{c} \tag{20}$$

The orbit $\vec{x}$ is now restricted to a range of m monitors and n correctors. Apart from the constraints (19) and (20) a further condition is that the correction should be closed.

The effect of the first $n-2$ correctors found by the algorithm on the orbit outside the short length region can be calculated and substituted by a single "virtual corrector" at a virtual position and phase. The effect of this "corrector" can be considered as a distortion to the outside orbit region and can be locally compensated by two additional correctors to ensure a closure. Since these two correctors, $n-1$ and n, are used for the closure, their strengths depend on the first $n-2$ correctors. Therefore, they cannot be varied independently and are not considered to find the best correction. However, their effect on the correction has to be included and the closure condition enters as a constraint into the minimization. After the correction has converged their strengths is calculated explicitly and they are added to the list of used correctors. The mathematics of this strategy is quite involved and a detailed discussion can be found in [8].

The minimization procedure can be described approximately as following:

1. Find a starting point $\vec{y}^{\,0}$ compatible with the constraints (19) and (20)

2. For the kth step: find a search direction as $d^k = -\nabla S(\vec{y}^{\,k-1})$
   (Direction of steepest gradient)

3. Without constraints: $d^{k,proj} = d^k$

   With constraints $\mathrm{B} \cdot \vec{y}^{\,k} = \vec{b}$:

   $d^{k,proj} = d^k - B^t\lambda \quad$ where $\quad \lambda = (BB^t)^{-1}Bd^k$
   (Gradient Projection, project onto subspace with fulfilled constraints)

4. Calculate maximum step length allowed by (20) and scan along the search direction for the minimum and keep it for the next step

5. Calculate $\vec{y}^{\,k+1}$ and stop if a satisfactory solution is found, otherwise go to 2. step
   A solution is found if $d^k = 0$ or $d^{k,proj} = 0$ and $\lambda \geq 0$

Unlike the method involving the MICADO algorithm, this method uses all available correctors inside the defined region. The convergence is usually good and the closure is naturally better than with the first method. Up to a total of about 10 correctors used it is faster than MICADO to find the desired correction.

One special feature of this method is that the norm for the minimization can be chosen as l_2 (least squares) or l_1. It was considered useful to have this possibility available for short length corrections.

4.2 Local bumps

The simplest form of a "short length correction" is the application of a closed bump. It is usually used for very local corrections or wanted orbit distortions and the possibility to make bumps with 2, 3 or 4 correctors is provided in COCU as well as a so-called sliding bump facility.

4.2.1 SINGLE BUMPS

A single bump can be applied at any position of the machine provided the necessary correctors are available. For a bump with only two correctors, a phase relationship between the correctors must be maintained (i.e. $\Delta\mu$ is a multiple of π), otherwise it is not possible to close the bump. Such a closure is always possible if three independent corrector magnets are used and the specified bump amplitude can always be obtained although the orbit may be strongly distorted in the neighborhood of the desired bump. In order to establish a bump with a specific amplitude *and* angle at any position, it is necessary to use four independent magnets. All these possibilities are foreseen and a detailed description can be found in [9].

4.2.2 SLIDING BUMPS

It is in principle possible to use bump algorithms to correct the whole closed orbit. Such schemes have been implemented in the former closed orbit correction procedure for the SPS.

In such a scheme, the deviations of the measured orbit from the ideal one are corrected one by one and the stengths of the used correctors are summed up. Each corrector is then shared by two bumps (two corrector π-bumps) or three bumps (three corrector bumps). When the orbit is a superposition of many free oscillations each caused by a single defect, the strengths of the correctors cancel in such a way that the remaining corrector strengths compensate exactly these defects. The method should therefore be equivalent to a global correction method such as MICADO or PSINOM (see e.g. [1, 2, 7]) provided no monitors or correctors are missing and the machine is "reasonably" regular. This has been tested experimentally on the SPS (with sliding three corrector bumps) and perfect agreement has been found.

The advantage of the method is that for an accelerator like the SPS the closed orbit can be (in theory) corrected to the ideal orbit with very little effort and almost automatically. Unfortunately the method is rather sensitive to wrong or missing beam position measurements and missing correctors.

5 Data preparation and conditioning

Before any of the described algorithms can be applied, the input data has to be prepared for the correction. The orbit measurements have to be read as well as the theoretical Twiss parameters. Consistency checks are necessary to avoid corrections with incompatible data.

The response matrix must be set up from the Twiss parameters according to eq (2), and if necessary, pre-conditioned to avoid numerical problems when the correction algorithms are applied. Systematic effects on the orbit, such as a momentum error, have to be identified and removed.

Special care has to be taken for the treatment of bad measurements and faulty equipment. Monitors and correctors which are faulty must be recognized and treated accordingly.

5.1 Off momentum particles

The energy error is taken into account by first determining this error from the measured orbit and then using a renormalized orbit in the correction algorithm. The energy error $\Delta p/p$ can be determined by minimizing the expression

$$S = \sum_{i=1}^{m}(r_i - D_i \cdot \frac{\Delta p}{p})^2 \tag{21}$$

where r_i is the radial beam position and D_i the theoretical dispersion function at the i^{th} monitor.

This is important if the beam is not centred on the nominal orbit since the correction would try to move the beam to the central orbit otherwise. For relativistic particles the mean radial position is determined by the RF frequency and dipole fields cannot change this mean value. Any attempt to correct this with dipoles would lead to wrong results. This option can be switched off, if desired.

5.2 Treatment of faulty monitors and correctors

The recognition and treatment of bad measurements is of vital importance for a successful orbit correction. Such bad measurements can be either localized missing or wrong single measurements or an extended region of the machine cannot be measured correctly.

Using false measurements in the orbit correction procedure can lead to severe problems and undesirable features, such as:

- Wrong or unsatisfactory corrections
- Using correctors far away from real distortion
- Introduction of unwanted corrections such as bumps etc.
- Numerical problems leading to an abortion of the correction

Hardware flags set by the low level software indicate faulty stations and some simple checks are built into the application software, such as cuts on the measurements exceeding several standard deviations of the measurement.

The philosophy used in COCU is that any monitor or corrector declared as bad or faulty ceases to exist for the whole procedure. The appropriate actions have to be performed, such as compressing the vectors $\vec{x}_m$ and $\vec{y}$ and the response matrix.

Furthermore, a missing monitor can lead to numerical problems: two columns in the response matrix corresponding to two different correctors can become linearly dependent if no valid measurement between them is available and their phase advance is a multiple of $n \cdot \pi$.

This is particularly important when the phase advance per cell is 90^0 as already one missing monitor leads to this problem. It is therefore necessary to remove the offending correctors from the list. This is done by different methods of which some are designed for speed and safety and others for an optimum correction. The principle of these methods is briefly discussed in the next section and a detailed discussion of this problem can be found in [12].

5.2.1 Matrix conditioning

Finding linearly dependent columns

The first method is a reconditing of the matrix itself by identifying linearly dependent columns and removing the associated correctors [13]. This works reliably but requires a relatively large amount of computing time which can dominate the time needed for the whole correction process. As this procedure would be required for every orbit correction, even if only one corrector is desired, this method can become too heavy and time consuming for regular use.

Ignoring redundant correctors

A second method which is fast and safe was required where the possible problems were anticipated. The method was to identify a situation where physically two correctors were separated by approximately 180^0 and no valid measurement in between would make them independent. However, to ensure a fast and safe correction, more correctors than strictly necessary were removed. In some cases this has caused problems to obtain the optimum orbit correction.

5.2.2 Singular value decomposition

A very rigorous method to avoid numerical problems with a singular response matrix, a Singular Value Decomposition (SVD) can be used to determine the real rank of the matrix, identify the singular values and remove the corresponding correctors. With the presently available computing power this may seem feasible and was therefore tried.

Furthermore, a proper and rigorous conditioning of the response matrix is a prerequisite for the use of the PSINOM algorithm.

The principle

The principle of a SVD is the following: for any rectangular matrix A[m x n] with rank r, there exist unitary matrices U[m x m] and V[n x n] such that

$$A = U \cdot \Sigma \cdot V \qquad \mathit{with} \qquad \Sigma = \begin{bmatrix} \sigma_1 & 0 & . & 0 & 0 \\ 0 & \sigma_2 & . & 0 & 0 \\ . & . & . & . & . \\ 0 & 0 & . & \sigma_r & 0 \\ 0 & 0 & . & 0 & 0 \\ 0 & 0 & . & 0 & 0 \\ 0 & 0 & . & 0 & 0 \end{bmatrix}$$

I.e. Σ is an [m x n] matrix whose elements are all zero except for the diagonal elements $\sigma_1 ... \sigma_r$, which are the singular values of the matrix A and are ordered according to their values: $\sigma_1 \geq \sigma_2 \geq ... \geq \sigma_r$.

How can we use it ?

It can be shown, that the singular values σ_i provide a measure for the "singularity" of the matrix A. In particular, the smallest singular value σ_r is a measure how close A is to be rank deficient [15].

The procedure is now as follow: find the singular values of the response matrix A. Identify the correctors which correspond to the zero values or too small σ_k and remove them from the available corrector set. The algorithm used in COCU is from Golub and Reinsch in [16].

5.3 Postprocessing

5.3.1 Maximum correctors strength in MICADO

In the MICADO algorithm it is not easy to implement constraints such as a limited corrector strength. In particular at the SPS, the maximum strength is relatively weak when it is run at high energies. Often the calculated correction cannot be applied due to the lack of strength. We have implemented a strategy to apply a correction under these circumstances. This possible technique is to fix the corrector concerned at its maximum value and to recalculate the remaining correctors [1]. For further corrections this corrector remains at its maximum value. This method was tried during the runs of the SPS proton-antiproton collider and we achieved convergence in most cases although this method is not a safe strategy to deal with this problem.

5.3.2 Orbit smoothing and bump reduction

One feature which is particularly disturbing is the appearance or two magnet bumps in the ring. Two effects are responsible for this problem. When MICADO is run for many iterations it may happen that a corrector is chosen which is one of a pair of correctors which could form a local bump to reduce a very local orbit distortion, whether it is real or not. In the following iteration MICADO will usually find the appropriate "partner corrector" and often sets both to very large strengths.

Another mechanism for the development of large corrector strengths is an accumulation process. If the closed orbit has changed due to a small magnet move or other drifts, MICADO can correct it by applying an appropriate correction. If the origin of the distortion is reversed, e.g. the magnet moves back into the original position, there is a chance that the previous correction is *not* reversed, but compensated by a correction with a corrector $n \cdot \pi$ away in betatron phase. In that case the correction has the *same* sign as the original one. Should the magnet drift happen again, in the worst case a correction is found which now *adds* to the first correction and so on. The strengths in the two correctors can build up quickly and appear as a two magnet bump which develops between the correctors. This phenomenon has been observed at LEP especially for the correctors close to the final focusing quadrupoles where the phase advance is very close to π and the quadrupoles themselves were drifting back and forth occasionally. Furthermore, a bump develops between the two correctors around the interaction point which is obviously very undesirable.

These types of bumps can be recognized and eliminated by a postprocessing of the correction. The corrector strengths are searched for the appearance of such bumps and an attempt is made to eliminate them by reducing the strengths by an equal amount for both correctors. For a phase advance between the two correctors not exactly 180^0 the resulting non-closure or leakage must be compensated with additional correctors. This is done using a technique very similar to that used for the closure of the GRAPE algorithm [14].

To avoid an accumulation into a pair of correctors, a special technique is applied at LEP for the focusing quadrupoles: only one of the two correctors is used in the correction algorithm and the strength calculated is distributed equally to both correctors before the hardware is loaded. This can avoid the accumulation process described above and although this is not performed inside COCU it is mentioned for completeness as it may be a technique of general interest.

5.3.3 Precision of corrector strength

It must be mentioned that not only the precision of the measurement, but also the precision of the corrector strength will influence the result of a correction and even may require special care in the algorithm.

One particular example is the short length correction with the GRAPE method. In principle, the correction is exactly closed, but after the correction is applied a small oscillation may be observed which is the result of a small non-closure caused by the finite precision of the corrector strengths.

When the strengths of the closing correctors are calculated, the full precision of the calculated strengths are used which is much better than the precision of the actual corrector in the machine. The actually used corrector strengths are therefore truncated and this is normally not taken into account in the calculation of the closing corrector strengths. If the number of correctors used in GRAPE substantially exceeds the number of closing correctors (i.e. 2), this truncation can lead to a small but non-negligible non-closure of the correction [8]. It is therefore possible in COCU to specify the precision of the correctors and the closure is calculated taking this into account. This has been verified by simulation and the tests in LEP fully confirmed this strategy: the non-closure was reduced by more than a factor two compared to the standard case where no special measures were taken [8].

6 Summary

The algorithms and procedures used in the closed orbit correction package COCU (Closed Orbit Correction Utilities) were described. Emphasis was put on a presentation of the basic principles and not the underlying mathematics or the implementation. The advantages and disadvantages of the various methods were discussed and the data preparation in COCU was described.

References

[1] W. Herr, T. Limberg, J. Miles, R. Schmidt; *A new Closed Orbit Correction procedure for the CERN-SPS*;
1989 Particle Accelerator Conference, Chicago, USA, March 20-23

[2] D. Brandt, W. Herr, J. Miles and R. Schmidt; *A New Closed Orbit Correction Procedure for the CERN SPS and LEP*;
Nucl. Instr. and Meth., **A293** (1990) 305. and CERN/SPS/89-48 (AMS).

[3] G. Crockford, W. Herr and V. Paris; *Application user guide for SPS and LEP closed orbit correction*; SL/Note 94-25 (OP) (1994).

[4] M. Albert, G. Crockford, W. Herr and V. Paris; *COCU user guide, Version 10.30*; (1995).

[5] Proceedings of the 5th workshop on LEP performance, CERN SL/95-08 (DI).

[6] S. Myers; *Conclusions of the 5th LEP performance workshop*; CERN SL/95-10 (DI).

[7] B. Autin and Y. Marti; *Closed Orbit Correction of A.G. Machines using a limited Number of Magnets*; CERN ISR MA/73-17(1973)

[8] T. Lundberg; Diploma Thesis, University Linkøping; *A New Short Length Orbit Correction Method*; SL Note/91-20 (AP).

[9] G. Crockford and W. Herr; *Bump package reference and user guide*, CERN SL/93-40 (AP).

[10] S.I. Gass; *Linear Programming*, McGraw Hill, New York, 1958.

[11] T. Risselada, Private communication.

[12] W. Herr; *How do we deal with missing monitors and redundant correctors in COCU ?*; SL Note/95-29 (AP).

[13] T. Lundberg; (1991) unpublished.

[14] T. Lundberg; *Suppression of Large Corrector Strength Produced by MICADO*; SL Note/91-10 (AP).

[15] S. Barnett; "Matrices, Methods and Applications", Clarendon Press, Oxford (1990)

[16] J.H. Wilkinson and C. Reinsch; "Handbook for Automatic Computation", Vol.2, Springer Verlag.